OECD Principles and Guidelines for Access to Research Data from Public Funding

OECD

OECD Principles and Guidelines for Access to Research Data from Public Funding

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

The OECD is a unique forum where the governments of 30 democracies work together to address the economic, social and environmental challenges of globalisation. The OECD is also at the forefront of efforts to understand and to help governments respond to new developments and concerns, such as corporate governance, the information economy and the challenges of an ageing population. The Organisation provides a setting where governments can compare policy experiences, seek answers to common problems, identify good practice and work to co-ordinate domestic and international policies.

The OECD member countries are: Australia, Austria, Belgium, Canada, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Japan, Korea, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Slovak Republic, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States. The Commission of the European Communities takes part in the work of the OECD.

OECD Publishing disseminates widely the results of the Organisation's statistics gathering and research on economic, social and environmental issues, as well as the conventions, guidelines and standards agreed by its members.

This work is published on the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of the Organisation or of the governments of its member countries.

© OECD 2007

No reproduction, copy, transmission or translation of this publication may be made without written permission. Applications should be sent to OECD Publishing *rights@oecd.org* or by fax 33 1 45 24 99 30. Permission to photocopy a portion of this work should be addressed to the Centre français d'exploitation du droit de copie (CFC), 20, rue des Grands-Augustins, 75006 Paris, France, fax 33 1 46 34 67 19, *contact@cfcopies.com* or (for US only) to Copyright Clearance Center (CCC), 222 Rosewood Drive Danvers, MA 01923, USA, fax 1 978 646 8600, *info@copyright.com*.

Foreword

Innovative scientific research has a crucial role in addressing global challenges — ranging from health care and climate change to renewable energy and natural resources management. The speed and depth of this research depends on fostering collaborative exchanges between different communities and assuring its widest dissemination. The exchange of ideas, knowledge and data emerging is fundamental for human progress and is part of the core of OECD values. Thus, I am very pleased that the OECD has taken the lead in developing principles and standards to facilitate access to research data generated with public funding.

The rapid development in computing technology and the Internet have opened up new applications for the basic *sources* of research — the base material of *research data* — which has given a major impetus to scientific work in recent years. Databases are rapidly becoming an essential part of the infrastructure of the global science system. The international Human Genome Project is but one good example of a large-scale endeavour in which openly accessible information is being used successfully by many different users, all over the world, for a great variety of purposes.

Besides, access to research data increases the returns from public investment in this area; reinforces open scientific inquiry; encourages diversity of studies and opinion; promotes new areas of work and enables the exploration of topics not envisioned by the initial investigators.

Science and Technology Ministers called on the OECD in 2004 to develop a set of guidelines based on commonly agreed principles to facilitate cost-effective access to digital research data from public funding. The attached Principles and Guidelines are the outcome of this request. They are intended to assist all actors involved when trying to improve the international sharing of, and access to, research data. Now that we have developed this useful instrument, I call upon political and scientific leaders to adopt it thoroughly. Its use will undoubtedly facilitate the scientific endeavour and therefore contribute to the betterment of society.

Angel Gurria
OECD Secretary-General

Table of Contents

Background 7

Preface 9

Principles and Guidelines for Access to Research Data from Public Funding 13

I. Objectives 13

II. Scope and definitions 13

Research data 13

Research data from public funding 14

Access arrangements 14

III. Principles 15

A. Openness 15

B. Flexibility 15

C. Transparency 15

D. Legal conformity 16

E. Protection of intellectual property 16

F. Formal responsibility 17

G. Professionalism 18

H. Interoperability 19

I. Quality 19

J. Security 20

K. Efficiency 21

L. Accountability 21

M. Sustainability 22

Background

In January 2004, Ministers of science and technology of OECD countries met in Paris and discussed the need for international guidelines on access to research data. At that meeting, the governments of the 30 OECD countries and of China, Israel, Russia and South Africa adopted a *Declaration on Access to Research Data from Public Funding*. In this declaration, they recognised the importance of access to research data and invited the OECD "to develop a set of OECD guidelines based on commonly agreed principles to facilitate optimal cost-effective access to digital research data from public funding to be endorsed by the OECD Council at a later stage".

This request was taken up by OECD's Committee for Scientific and Technological Policy, which launched a project by asking a group of experts to develop a set of principles and guidelines. The experts drafted a first set of principles and guidelines and engaged in several rounds of consultation with research institutions and policy making bodies in the OECD member countries to achieve a consensus. A workshop involving key stakeholders was held in Paris in February 2006 which also contributed to this process. The work leading up to the final draft revealed that international frameworks to facilitate access were still lacking in the member countries, but also that improved access was generally seen as benefiting the advancement of research, boosting its quality and facilitate cross-disciplinary research co-operation. Stakeholders considered that international guidelines would be useful in giving guidance to institutions in need of policies and in enhancing international co-operation in research. The principles and guidelines that resulted from this extensive consultation process were approved by the OECD's Committee for Scientific and Technological Policy in October 2006. The Principles and Guidelines were attached to an OECD Recommendation and endorsed by the OECD Council on 14 December 2006.

OECD Recommendations set out collective and precise standards or objectives which the member countries are expected to implement. A "Recommendation" is a legal instrument of the OECD that is not legally binding but through a long standing practice of the member countries, is considered to have a great moral force. Recommendations of the OECD are adopted when member governments are prepared to make a political commitment to implement the principles (and/or guidelines) set out therein. This type of instrument is often referred to as "soft law".

Although some flexibility in the drafting of the standard is possible, the flexibility should not be so wide as to allow individual member countries to modify the standard or object, consequently defeating the idea that there is a commitment on their part. That said, OECD Recommendations often leave significant flexibility to member countries as regards the means through which they implement them to take into account differences in legal, cultural, economic and social contexts. Thus, in some countries, implementation could be achieved through regulatory measures while other will choose, for example to resort to concerted action with national stakeholders.

Recommendations are considered to be vehicles for change, and OECD member countries need not, on the day of adoption, already be in conformity. What is expected is that they will seriously work towards attaining the standard or objective within a reasonable time frame considering the extent of difficulty in closing the gap in each member country.

The OECD Recommendation states that member countries should take into consideration the Principles and Guidelines on Access to Research Data from Public Funding set out in the Annex to the Recommendation, as appropriate for each member country, to develop policies and good practices related to the accessibility, use and management of research data. It also instructs OECD's Committee for Scientific and Technological Policy to review the implementation of this Recommendation as necessary; and to review the Principles and Guidelines on Access to Research Data from Public Funding when appropriate, to take into account advances in technology and research practices, with the intention of further fostering international co-operation.

The attached *Principles and Guidelines* are meant to apply to research data that are gathered using public funds for the purposes of producing publicly accessible knowledge. The nature of "public funding" of research varies significantly from one country to the next, as do existing data access policies and practices at the national, disciplinary and institutional levels. These differences call for a flexible approach in developing data access arrangements. The balance between the costs of improved access to research data and the benefits that result from such access will need to be judged by individual national governments and their research communities.

Preface

Increasing the return on public investments in scientific research

The public science systems of OECD member countries are based on the principle of openness and the free exchange of ideas, information and knowledge. New information and communication technologies (ICTs), now in widespread use throughout all research disciplines, have greatly aided this system of free exchange and have opened up new avenues for collaboration and sharing. The progress of science, however, depends on more than just technologies. Research policies, practices, support systems and cultural values all affect the nature of new discoveries, the rate at which they are made, and the degree to which they are made accessible and used.

The power of computers and the Internet has created new fields of application for not only the results of research, but the *sources* of research: the base material of *research data*. Moreover, research data, in digital form, are increasingly being used in research endeavours beyond the original project for which they were gathered, in other research fields and in industry. Administrative data from the institutions of OECD member countries, such as employment information, are now used extensively in the social sciences, as well as in policy making. Data from public health organisations play a growing role in the advancement of life sciences. Similarly, geo-spatial data collected by many different government organisations are essential for environmental and other types of research. The list goes on.

Scientific databases are rapidly becoming a crucial part of the infra-structure of the global science system. The international *Human Genome Project* is but one good example of a large-scale research endeavour in which an openly accessible data repository is being used successfully by many different researchers, all over the world, for different purposes and in different contexts. Many other examples, involving research undertakings both large and small, are readily available.

Effective access to research data, in a responsible and efficient manner, is required to take full advantage of the new opportunities and benefits offered by ICTs. Accessibility to research data has become an important condition in:

- The good stewardship of the public investment in factual information;
- The creation of strong value chains of innovation;
- The enhancement of value from international co-operation.

More specifically, improved access to, and sharing of, data:

- Reinforces open scientific inquiry;
- Encourages diversity of analysis and opinion;
- Promotes new research;
- Makes possible the testing of new or alternative hypotheses and methods of analysis;
- Supports studies on data collection methods and measurement;
- Facilitates the education of new researchers;
- Enables the exploration of topics not envisioned by the initial investigators;
- Permits the creation of new data sets when data from multiple sources are combined.

Sharing and open access to publicly funded research data not only helps to maximise the research potential of new digital technologies and networks, but provides greater returns from the public investment in research.

Throughout OECD member countries, continuously growing quantities of data are collected by publicly funded researchers and research institutions. This rapidly expanding body of research data represents both a massive investment of public funds and a potential source of the knowledge needed to address the myriad challenges facing humanity.

To promote improved scientific and social return on the public investments in research data, OECD member countries have established a variety of laws, policies and practices concerning access to research data at the national level. In this context, international guidelines would be an important contribution to fostering the global exchange and use of research data.

These *Principles and Guidelines* are meant to apply to research data that are gathered using public funds for the purposes of producing publicly accessible knowledge. The nature of "public funding" of research varies significantly from one country to the next, as do existing data access policies and practices at the national, disciplinary and institutional levels. These differences call for a flexible approach to data access and recognition that one size does not fit all. Moreover, the balance between the costs of improved

access to research data and the benefits that result from such access will need to be judged by individual national governments and their research communities.

Whatever differences there may be between practices of, and policies on, data sharing, and whatever legitimate restrictions may be put on data access, practically all research could benefit from more systematic sharing. As the authors of the US National Research Council study, *Bits of Power,* pointed out:

> *The value of data lies in their use. Full and open access to scientific data should be adopted as the international norm for the exchange of scientific data derived from publicly funded research.*

The specific aims and objectives of these *Principles and Guidelines* are to:

- Promote a culture of openness and sharing of research data among the public research communities within member countries and beyond;
- Stimulate the exchange of good practices in data access and sharing;
- Raise awareness about the potential costs and benefits of restrictions and limitations on access to and the sharing of research data from public funding;
- Highlight the need to consider data access and sharing regulations and practices in the formation of member countries' science policies and programmes;
- Provide a commonly agreed upon framework of operational principles for the establishment of research data access arrangements in member countries;
- Offer recommendations to member countries on how to improve the international research data sharing and distribution environment.

The *Principles and Guidelines* contained in this document should assist governments, research support and funding organisations, research institutions and researchers themselves in dealing with the barriers and challenges in improving the international sharing of, and access to, research data. These *Principles and Guidelines* should be considered in light of, and applied to, the following major issues inherent in providing data access:

- Technological issues: access to research data, and their optimum exploitation, requires appropriately designed technological infrastructure, broad international agreement on interoperability, and effective data quality controls.

- Institutional and managerial issues: while increased accessibility is important to all science communities, the diversity of the scientific enterprise suggests that a variety of institutional models and tailored data management approaches are most effective in meeting the needs of researchers.

- Financial and budgetary issues: scientific data infrastructure requires continued and dedicated budgetary planning and appropriate financial support. The use of research data will not be maximised if access, management, and preservation costs are an add-on or after-thought in research projects. It is important to note, however, that the cost of storing and managing data has decreased dramatically in recent years, and lack of knowledge about such changes can, in itself, be a barrier to advancement.

- Legal and policy issues: national laws and international agreements, particularly in areas such as intellectual property rights and the protection of privacy, directly affect data access and sharing practices, and must be fully taken into account in the design of data access arrangements.

- Cultural and behavioural issues: appropriate educational and reward structures are a necessary component for promoting data access and sharing practices. These considerations apply to those who fund, produce, manage, and use research data.

In working towards better access to research data in the context of these *Principles and Guidelines*, member countries will need to determine the appropriate balance between the costs of improved access to this data and the benefits that result from such access. The efforts to improve access, of course, need to be carried out within existing financial limitations.

PRINCIPLES AND GUIDELINES FOR ACCESS TO RESEARCH DATA FROM PUBLIC FUNDING

I. Objectives

These *Principles and Guidelines for Access to Research Data from Public Funding* (hereafter the "*Principles and Guidelines*") provide broad policy recommendations to the governmental science policy and funding bodies of member countries on access to research data from public funding. They are intended to promote data access and sharing among researchers, research institutions, and national research agencies, while at the same time, recognising and taking into account, the various national laws, research policies and organisational structures of member countries.

The ultimate goal of these *Principles and Guidelines* is to improve the efficiency and effectiveness of the global science system. They are not intended to hinder its development with onerous obligations and regulations or impose new costs on national science systems.

II. Scope and definitions

These *Principles and Guidelines* are meant to apply to research data, whether already in existence or yet to be produced, that are supported by public funds for the purposes of developing publicly accessible scientific research and knowledge. The *Principles and Guidelines* are not intended to apply to research data gathered for the purpose of commercialisation of research outcomes, or to research data that are the property of a private sector entity. Access to such data is subject to a range of considerations that are beyond the scope of this document. Moreover, in some instances, access to or use of data may be restricted to safeguard the privacy of individuals, protect confidentiality, proprietary results or national security.

Research data

In the context of these *Principles and Guidelines*, "research data" are defined as factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings. A research data set constitutes a systematic, partial representation of the subject being investigated.

This term does not cover the following: laboratory notebooks, preliminary analyses, and drafts of scientific papers, plans for future research, peer reviews, or personal communications with colleagues or physical objects (*e.g.* laboratory samples, strains of bacteria and test animals such as mice). Access to all of these products or outcomes of research is governed by different considerations than those dealt with here.

These *Principles and Guidelines* are principally aimed at research data in digital, computer-readable format. It is indeed in this format that the greatest potential lies for improvements in the efficient distribution of data and their application to research because the marginal costs of transmitting data through the Internet are close to zero. These *Principles and Guidelines* could also apply to analogue research data in situations where the marginal costs of giving access to such data can be kept reasonably low.

Research data from public funding

Research data from public funding is defined as the research data obtained from research conducted by government agencies or departments, or conducted using public funds provided by any level of government. Given that the nature of "public funding" of research varies significantly from one country to the next, these *Principles and Guidelines* recognise that such differences call for a flexible approach to improved access to research data.

Access arrangements

Access arrangements are defined as the regulatory, policy and procedural framework established by research institutions, research funding agencies and other partners involved, to determine the conditions of access to and use of research data.

III. Principles

A. Openness

Openness means access on equal terms for the international research community at the lowest possible cost, preferably at no more than the marginal cost of dissemination. Open access to research data from public funding should be easy, timely, user-friendly and preferably Internet-based.

B. Flexibility

Flexibility requires taking into account the rapid and often unpredictable changes in information technologies, the characteristics of each research field and the diversity of research systems, legal systems and cultures of each member country. Specific national, social, economic and regulatory implications should be considered when organisations develop research data access arrangements, and when governments develop policies to promote data access and review the implementation of these *Principles and Guidelines*.

C. Transparency

Information on research data and data-producing organisations, documentation on the data and specifications of conditions attached to the use of these data should be internationally available in a transparent way, ideally through the Internet. Lack of visibility of existing research data resources and future data collection poses serious obstacles to access.

Factors to consider in ensuring transparency include:

- Information on data-producing organisations and their holdings, documentation on available data sets and conditions of use should be easy to find on the Internet.
- Research organisations and government research agencies should actively disseminate information on research data policies to individual researchers, academic associations, universities and other stakeholders in the publicly funded research process.
- Whenever relevant, all members of the various research communities should assist in establishing agreements on standards for cataloguing data. The application of existing standards should be considered, whenever appropriate, in order to avoid placing additional burdens on research resources and work loads of researchers and their institutions.

- Information on data management and access conditions should be communicated among data archives and data producing institutions, so that best practices can be shared.

D. Legal conformity

Data access arrangements should respect the legal rights and legitimate interests of all stakeholders in the public research enterprise.

Access to, and use of, certain research data will necessarily be limited by various types of legal requirements, which may include restrictions for reasons of:

- National security: data pertaining to intelligence, military activities, or political decision making may be classified and therefore subject to restricted access.
- Privacy and confidentiality: data on human subjects and other personal data are subject to restricted access under national laws and policies to protect confidentiality and privacy. However, anonymisation or confidentiality procedures that ensure a satisfactory level of confidentiality should be considered by custodians of such data to preserve as much data utility as possible for researchers.
- Trade secrets and intellectual property rights: data on, or from, businesses or other parties that contain confidential information may not be accessible for research.
- Protection of rare, threatened or endangered species: in certain instances there may be legitimate reasons to restrict access to data on the location of biological resources for the sake of conservation.
- Legal process: data under consideration in legal actions (sub judice) may not be accessible.

Subscribing to professional codes of conduct may facilitate meeting legal requirements.

E. Protection of intellectual property

Data access arrangements should consider the applicability of copyright or of other intellectual property laws that may be relevant to publicly funded research databases. Factors to consider include:

- As public/private partnerships in the funding of research and related data production are increasing, balanced public/private arrangements should facilitate broad access to research data where appropriate. The fact that there is private sector involvement in the data collection

should not, in itself, be used as a reason to restrict access to the data. Consideration should be given to measures that promote non-commercial access and use while protecting commercial interests, such as delayed or partial release of such data, or the voluntary adoption of licensing mechanisms. Such measures can allow the primary participants to fully exploit the research data without unnecessarily shutting off access.

- In those jurisdictions in which government research data and information are protected by intellectual property rights, the holders of these rights should nevertheless facilitate access to such data particularly for public research or other public-interest purposes.

F. Formal responsibility

Access arrangements should promote explicit, formal institutional practices, such as the development of rules and regulations, regarding the responsibilities of the various parties involved in data-related activities. These practices should pertain to authorship, producer credits, ownership, dissemination, usage restrictions, financial arrangements, ethical rules, licensing terms, liability, and sustainable archiving.

Access arrangements, whether at the governmental or institutional levels, should be developed in consultation with representatives of all directly affected parties. In collaborative research programmes or projects, and especially in international scientific co-operation or in research projects based on public/private partnerships where there are differences in regulatory frameworks, the parties involved should negotiate research data sharing arrangements as early as possible in the life of the research project, ideally at the initial proposal stage. This will help ensure that adequate and timely consideration will be given to issues such as the allocation of resources for sharing and sustainable preservation of research data, differences in national intellectual property laws, limitations due to national security, and the protection of privacy and confidentiality.

Access arrangements also should be responsive to factors such as the characteristics of the data, their potential value for research purposes, the level of data processing (raw versus partially processed versus final), whether they are homogeneous data from a facility instrument or sensor versus heterogeneous field data collected by single researchers, data on human subjects or physical parameters, and whether the data are generated directly by a government entity or as a result of government funding. These variations in the origin or type of data should be taken into consideration when establishing data access arrangements.

Further, consideration should be given to the following:

- Many of the problems related to access, dissemination and sharing of data result from the lack of explicit institutional agreements on the terms of access and use. With data management becoming ever more complex in certain areas of research, traditional informal arrangements between researchers may no longer be adequate and may need to be complemented by formally agreed practices and procedures.
- Responsibility for the various aspects of data access and management should be established in relevant documents, such as descriptions of the formal tasks of institutions, grant applications, research contracts, publication agreements, and licenses.
- Long-term sustainability of the infrastructure required for data access is particularly important. Research institutions and government organisations should take formal responsibility for ensuring that research data are effectively preserved, managed and made accessible in order that they can be put to efficient and appropriate use over the long term.

G. Professionalism

Institutional arrangements for the management of research data should be based on the relevant professional standards and values embodied in the codes of conduct of the scientific communities involved.

Factors to consider include:

- The use of codes of conduct for professional scientists and their communities could help simplify and reduce the regulatory burden placed on access.
- Mutual trust between researchers, and trust between researchers, their institutions and other organisations plays an important role in the establishment and maintenance of such codes of conduct.
- In current research practice, the initial data-producing researcher or institution is sometimes rewarded with temporary exclusive use of the data. The rules for such incentive arrangements should be developed and explicitly stated by the funding sources in co-operation with the affected research communities.

In certain areas of science, a lack of planning for and execution of the proper documentation and archiving of data sets is one of the key impedements to realising maximum value from the investment in research data. Project and program planning activities, at all levels, should expressly acknowledge data issues at the earliest stages to take into consideration funding and technical assistance for the essential organisation and curation of those data sets. Attention should be paid to incentives and the development of professional expertise in all areas of research data management.

H. Interoperability

Technological and semantic interoperability is a key consideration in enabling and promoting international and interdisciplinary access to and use of research data. Access arrangements, should pay due attention to the relevant international data documentation standards. member countries and research institutions should co-operate with international organisations charged with developing new standards.

Although science is becoming a highly globalised endeavour, incompatibility of technical and procedural standards can be the most serious barrier to multiple uses of data sets.

Factors that should be considered include:

- The standards employed should be explicitly mentioned as this is the first requirement for interoperability.
- Adoption of the practices of disciplines most advanced in this respect should be promoted, in particular by the international professional organisations dealing with science and the collection and preservation of data for research and technological purposes.
- The work of organisations engaged in setting more general information and communication technology standards should also be considered.

I. Quality

The value and utility of research data depends, to a large extent, on the quality of the data itself. Data managers, and data collection organisations, should pay particular attention to ensuring compliance with explicit quality standards. Where such standards do not yet exist, institutions and research associations should engage with their research community on their development. Although all areas of research can benefit from improved data quality, some require much more stringent standards than others. For this reason alone, universal data quality standards are not practical. Standards

should be developed in consultation with researchers to ensure that the level of quality and precision meets the needs of the various disciplines.

More specifically,

- Data access arrangements should describe good practices for methods, techniques and instruments employed in the collection, dissemination and accessible archiving of data to enable quality control by peer review and other means of safeguarding quality and authenticity.
- The origin of sources should be documented and specified in a verifiable way. Such documentation should be readily available to all who intend to use the data and incorporated into the metadata accompanying the data sets. Developing such metadata is important for enabling scientists to understand the exact implications of the data sets.
- Whenever possible, access to data sets should be linked with access to the original research materials, and copied data sets should be linked with originals, as this facilitates validation of the data and identification of errors within data sets.
- Research institutions and professional associations should develop appropriate practices with respect to the citations of data and the recording of citations in indexes, as these are important indicators of data quality.

J. Security

Specific attention should be devoted to supporting the use of techniques and instruments to guarantee the integrity and security of research data. With regard to guaranteeing the integrity of a data set, every effort should be made to ensure the completeness of data and absence of errors. With regard to security, the data, along with relevant meta-data and descriptions, should be protected against intentional or unintentional loss, destruction, modification and unauthorised access in conformity with explicit security protocols. Data sets and the equipment on which they are stored should be protected as well from environmental hazards such as heat, dust, electrical surges, magnetism, and electrostatic discharges.

K. Efficiency

One of the central goals of promoting data access and sharing is to improve the overall efficiency of publicly funded scientific research to avoid the expensive and unnecessary duplication of data collection efforts.

Consideration should be given to the following:

- Data access arrangements should promote further cost effectiveness within the global science system by describing good practices in data management and specialised support services.
- While publicly funded research data are subject to the default rule of openness under Principle A, this does not mean that all such data should be preserved permanently. The data archiving community should carry out cost-benefit assessments periodically and constantly develop and refine retention protocols to ensure that those data sets with the greatest potential utility are preserved and made accessible. Use of accepted retention protocols and thorough documentation of data should help to reduce unnecessary duplication of effort as well as to establish the necessary selectivity in preservation.
- Specialised support services, for example through collaboration with non-academic specialists on specific research projects or the engagement of data management specialist organisations, should be considered as a means to ensure the cost-effective production, use, management and archiving of research data.
- Insufficient incentives for researchers or database producers may lessen their efforts on data-related activities. The development of new reward structures and the adaptation of existing ones, including recognition of data management activities in tenure and promotion review, should be considered as a way to address this problem.

L. Accountability

The performance of data access arrangements should be subject to periodic evaluation by user groups, responsible institutions and research funding agencies. Although each party is likely to use somewhat different evaluation criteria, the sum total of the results should provide a comprehensive picture of the value of data and of data access regimes. Such evaluations should help to increase the support for open access among the scientific community and society at large.

The following should be considered in establishing evaluation criteria:

- Overall public investments in the production and management of research data.
- Management performance of data collection and archival agencies.
- Extent of re-use of existing data sets.
- Knowledge generated from the re-use of existing data.
- The use of targeted foresight exercises to determine the nature and scope of data preservation activities and the types of data most likely to be needed in the future.

Even if gaining clear insight into the cost, benefit and performance of data access arrangements will not be an easy task, those in charge of data access arrangements should put effort into showing the benefits of open data access to justify and help ensure sustained support from all levels of government.

M. Sustainability

Due consideration should be given to the sustainability of access to publicly funded research data as a key element of the research infrastructure. This means taking administrative responsibility for the measures to guarantee permanent access to data that have been determined to require long-term retention. This can be a difficult task, given that most research projects, and the public funding provided, have a limited duration, whereas ensuring access to the data produced is a long-term undertaking. Research funding agencies and research institutions, therefore, should consider the long-term preservation of data at the outset of each new project, and in particular, determine the most appropriate archival facilities for the data.

OECD PUBLICATIONS, 2, rue André-Pascal, 75775 PARIS CEDEX 16
PRINTED IN FRANCE
(00 2007 39 3 P) – No. 88180 2007

Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics

OCDE

Principes et lignes directrices pour l'accès aux données de la recherche financée sur fonds publics

OCDE

ORGANISATION DE COOPÉRATION ET DE DÉVELOPPEMENT ÉCONOMIQUES

ORGANISATION DE COOPÉRATION ET DE DÉVELOPPEMENT ÉCONOMIQUES

L'OCDE est un forum unique en son genre où les gouvernements de 30 démocraties œuvrent ensemble pour relever les défis économiques, sociaux et environnementaux que pose la mondialisation. L'OCDE est aussi à l'avant-garde des efforts entrepris pour comprendre les évolutions du monde actuel et les préoccupations qu'elles font naître. Elle aide les gouvernements à faire face à des situations nouvelles en examinant des thèmes tels que le gouvernement d'entreprise, l'économie de l'information et les défis posés par le vieillissement de la population. L'Organisation offre aux gouvernements un cadre leur permettant de comparer leurs expériences en matière de politiques, de chercher des réponses à des problèmes communs, d'identifier les bonnes pratiques et de travailler à la coordination des politiques nationales et internationales.

Les pays membres de l'OCDE sont : l'Allemagne, l'Australie, l'Autriche, la Belgique, le Canada, la Corée, le Danemark, l'Espagne, les États-Unis, la Finlande, la France, la Grèce, la Hongrie, l'Irlande, l'Islande, l'Italie, le Japon, le Luxembourg, le Mexique, la Norvège, la Nouvelle-Zélande, les Pays-Bas, la Pologne, le Portugal, la République slovaque, la République tchèque, le Royaume-Uni, la Suède, la Suisse et la Turquie. La Commission des Communautés européennes participe aux travaux de l'OCDE.

Les Éditions OCDE assurent une large diffusion aux travaux de l'Organisation. Ces derniers comprennent les résultats de l'activité de collecte de statistiques, les travaux de recherche menés sur des questions économiques, sociales et environnementales, ainsi que les conventions, les principes directeurs et les modèles développés par les pays membres.

Cet ouvrage est publié sous la responsabilité du Secrétaire général de l'OCDE. Les opinions et les interprétations exprimées ne reflètent pas nécessairement les vues de l'OCDE ou des gouvernements de ses pays membres.

© OCDE 2007

Toute reproduction, copie, transmission ou traduction de cette publication doit faire l'objet d'une autorisation écrite. Les demandes doivent être adressées aux Éditions OCDE *rights@oecd.org* ou par fax 33 1 45 24 99 30. Les demandes d'autorisation de photocopie partielle doivent être adressées au Centre français d'exploitation du droit de copie (CFC), 20, rue des Grands-Augustins, 75006 Paris, France, fax 33 1 46 34 67 19, *contact@cfcopies.com* ou (pour les États-Unis exclusivement) au Copyright Clearance Center (CCC), 222 Rosewood Drive Danvers, MA 01923, USA, fax 1 978 646 8600, *info@copyright.com*.

Avant-propos

La recherche scientifique au service de l'innovation est essentielle pour relever les défis qui se posent à l'échelle mondiale, qu'il s'agisse des soins de santé et du changement climatique, de l'énergie renouvelable ou de la gestion des ressources naturelles. Pour en accroître le rythme et l'intensité, il importe de promouvoir des échanges collaboratifs entre les différentes communautés et d'assurer la diffusion la plus large possible de ses résultats. L'échange des idées, des connaissances et des données qui se font jour est fondamental pour le progrès de l'humanité et fait partie des valeurs essentielles portées par l'OCDE. Je suis donc particulièrement heureux que l'OCDE ait ouvert la voie en élaborant des principes et des normes qui facilitent l'accès aux données de la recherche financée sur des fonds publics.

Le développement rapide de l'informatique et de l'Internet a permis l'émergence de nouvelles applications pour les *sources* mêmes de l'étude scientifique que constituent les *données de la recherche*, dopant ainsi puissamment l'activité scientifique de ces dernières années. Les bases de données deviennent rapidement un élément essentiel de l'infrastructure du système scientifique mondial. Le Projet international sur le génome humain est un bon exemple, parmi d'autres, d'une entreprise à grande échelle dans laquelle une information librement accessible est exploitée avec succès par un large éventail d'utilisateurs, partout dans le monde et à des fins les plus diverses.

Par ailleurs, l'accès aux données de la recherche accroît le retour sur l'investissement public dans ce domaine, il renforce la liberté de l'investigation scientifique, il encourage la diversité des études et opinions, il favorise de nouveaux domaines d'activité et permet l'exploration de thèmes qui n'avaient pas été envisagés par les chercheurs d'origine.

En 2004, les ministres de la science et de la technologie ont demandé à l'OCDE de définir un ensemble de lignes directrices à partir de principes établis d'un commun accord en vue de faciliter un accès au moindre coût aux données numériques de la recherche financée sur fonds publics. Les Principes et lignes directrices présentés dans ce document font suite à cette demande. Ils sont destinés à aider tous acteurs intéressés à améliorer le partage des données de la recherche et de l'accès à ces données à l'échelle internationale. Maintenant que nous disposons de ce précieux instrument, j'appelle les responsables politiques et scientifiques à l'adopter sans réserve. Son utilisation, j'en suis convaincu, facilitera l'activité scientifique et contribuera donc à l'amélioration de la société.

Angel Gurría
Secrétaire général de l'OCDE

Table des matières

Contexte 7

Préface 11

Principes et Lignes directrices pour l'accès aux données de la recherche financée sur fonds publics **17**

I. Objectifs 17

II. Champ d'application et définitions 17

Données de la recherche 18

Données de la recherche financée sur fonds publics 18

Dispositifs d'accès 18

III. Principes 19

A. Ouverture 19

B. Flexibilité 19

C. Transparence 19

D. Conformité au droit 20

E. Protection de la propriété intellectuelle 21

F. Responsabilité formelle 21

G. Professionnalisme 23

H. Interopérabilité 23

I. Qualité 24

J. Sécurité 25

K. Efficience 25

L. Responsabilité de rendre compte 26

M. Pérennité 27

Contexte

Les ministres de la science et de la technologie des pays membres de l'OCDE se sont réunis à Paris en janvier 2004 pour étudier la nécessité d'élaborer des lignes directrices internationales sur l'accès aux données de la recherche. Lors de cette réunion, les gouvernements des 30 pays membres de l'OCDE ainsi que ceux de l'Afrique du Sud, de la Chine, d'Israël et de la Russie ont adopté une *Déclaration ministérielle sur l'accès aux données de la recherche financée par des fonds publics*, dans laquelle ils reconnaissent l'importance de l'accès aux données de la recherche et invitent l'OCDE « à définir un ensemble de lignes directrices fondées sur un ensemble convenu de principes destiné à faciliter l'accès, dans des conditions optimales d'efficacité par rapport aux coûts, aux données numériques de la recherche, pour approbation par le Conseil de l'OCDE à un stade ultérieur ».

Le Comité de la politique scientifique et technologique de l'OCDE, qui s'est chargé de cette demande, a commencé par demander à un groupe d'experts d'élaborer un ensemble de principes et lignes directrices. Ces experts ont rédigé un premier ensemble de principes et lignes directrices et entrepris plusieurs cycles de consultation avec des institutions de recherche et des organes directeurs dans les pays membres de l'Organisation afin de parvenir à un consensus. Un atelier organisé à Paris en février 2006 avec des acteurs clés a également contribué à ce processus. Les travaux menés pour la version finale ont montré que les pays membres manquaient encore de cadres internationaux visant à faciliter l'accès aux données, ils ont également révélé qu'un meilleur accès était généralement perçu comme un atout pour faire avancer la recherche, pour renforcer sa qualité et faciliter la coopération scientifique pluridisciplinaire. Les parties intéressées ont estimé que des lignes directrices internationales seraient utiles pour donner des orientations aux institutions cherchant à définir une ligne de conduite ainsi que pour renforcer la coopération internationale dans le domaine de la recherche. Les principes et lignes directrices élaborés à la suite de ce vaste processus de consultation ont été approuvés par le Comité de la politique scientifique et technologique de l'OCDE en octobre 2006, puis joints à une Recommandation de l'OCDE et entérinés par le Conseil de l'OCDE le 14 décembre 2006.

Les Recommandations de l'OCDE définissent des normes ou des objectifs collectifs et précis que les pays membres doivent en principe mettre en œuvre. Une Recommandation de l'OCDE est un instrument de l'Organisation qui n'est pas juridiquement contraignant mais qui, selon un usage bien établi dans les pays membres, est considéré comme ayant un grand poids moral. Les Recommandations de l'OCDE sont adoptées lorsque les gouvernements des pays membres sont prêts à prendre l'engagement politique d'appliquer les principes (et/ou lignes directrices) ainsi définis. Ce type d'instrument relève de ce que l'on appelle souvent le « droit non contraignant ».

Bien qu'une certaine latitude dans la rédaction de la norme soit admise, elle ne doit pas être suffisamment grande pour laisser aux pays membres la possibilité de modifier la norme ou l'objectif, et compromettre ainsi l'idée d'engagement de leur part. Cela dit, les Recommandations de l'OCDE laissent souvent une grande marge de manœuvre aux pays membres pour ce qui est des moyens de mise en œuvre, afin de prendre en compte les différents contextes juridiques, culturels, économiques et sociaux. Certains pays peuvent ainsi les mettre en œuvre par voie réglementaire, alors que d'autres choisiront par exemple d'entreprendre une action concertée avec les parties intéressées au niveau national.

Les Recommandations sont considérées comme des vecteurs de changement, et les pays membres de l'OCDE ne sont pas obligés, le jour de l'adoption, d'en respecter déjà les termes. Le but est qu'ils s'emploient sérieusement à atteindre la norme ou l'objectif défini dans des délais raisonnables, compte tenu du degré de difficulté auquel chaque pays membre est confronté pour y parvenir.

Dans sa Recommandation, le Conseil de l'OCDE demande que les pays membres prennent en considération les Principes et lignes directrices pour l'accès aux données de la recherche financée sur fonds publics qui sont exposés à l'annexe de la Recommandation, selon le contexte de chaque pays membre, pour élaborer des politiques et bonnes pratiques en relation avec l'accessibilité, l'utilisation et la gestion des données de la recherche. Il charge également le Comité de la politique scientifique et technologique de l'OCDE d'examiner la mise en œuvre de ladite Recommandation en tant que de besoin, ainsi que de réexaminer les Principes et lignes directrices pour l'accès aux données de la recherche financée sur fonds publics et, le cas échéant, de prendre en compte les progrès de la technologie et des pratiques de la recherche, en vue de resserrer encore davantage la coopération internationale.

Les *Principes et Lignes directrices* présentés ci-après ont été élaborés pour s'appliquer aux données de la recherche recueillies à l'aide de fonds publics pour produire des connaissances publiquement accessibles. La nature du « financement public » de la recherche varie considérablement d'un pays à l'autre, tout comme les politiques et pratiques actuelles en matière d'accès aux données en fonction des pays, des disciplines et des institutions. Ces différences plaident en faveur d'une approche flexible de l'élaboration de dispositifs d'accès aux données. Il appartiendra à chaque gouvernement national et à sa communauté scientifique d'arbitrer entre les coûts d'un meilleur accès aux données de la recherche et les bénéfices pouvant en résulter.

Préface

Augmenter le rendement des investissements publics dans la recherche scientifique

Les systèmes scientifiques publics des pays membres de l'OCDE sont fondés sur le principe de l'ouverture et de la libre circulation des idées, de l'information et du savoir. Les nouvelles technologies de l'information et des communications (TIC), aujourd'hui omniprésentes dans toutes les disciplines de recherche, ont beaucoup favorisé ce système de libre circulation et ont ouvert de nouvelles voies pour la collaboration et le partage. Mais le progrès scientifique n'est pas tributaire des seules technologies. Les politiques et pratiques de la recherche, ainsi que les systèmes qui la soutiennent et les valeurs culturelles sont autant de facteurs qui influent sur la nature et la fréquence des découvertes et le degré avec lequel celles-ci sont rendues accessibles et utilisées.

La puissance des ordinateurs et l'internet ont donné naissance à de nouveaux champs d'application, non seulement pour les résultats de la recherche, mais pour ses sources mêmes, à savoir le matériau de base que constituent les données de la recherche. De plus, les données de la recherche sous forme numérique sont de plus en plus utilisées dans des travaux qui vont au delà du projet en vue duquel elles ont été recueillies à l'origine, ainsi que dans d'autres domaines de recherche et dans l'industrie. Les données administratives des institutions des pays membres de l'OCDE telles que les données sur l'emploi, sont aujourd'hui largement utilisées en sciences sociales ainsi que pour l'élaboration des politiques. Les données des organismes de santé publique jouent un rôle de plus en plus important dans l'avancement des sciences du vivant. De même, les données géospatiales recueillies par un grand nombre d'organismes gouvernementaux différents sont essentielles pour la recherche en environnement et autres domaines. Et la liste est longue.

Les bases de données scientifiques deviennent rapidement un élément crucial de l'infrastructure du système scientifique mondial. Le Projet international sur le génome humain n'est qu'un exemple parmi d'autres d'une entreprise de recherche à grande échelle dans laquelle une source de données librement accessible est largement utilisée par de nombreux chercheurs différents, partout dans le monde, à différentes fins et dans différents

contextes. On peut facilement trouver un grand nombre d'autres exemples impliquant des entreprises de recherche, à grande ou à petite échelle.

Un accès efficace aux données de la recherche, utilisé de façon responsable et efficiente, est nécessaire pour tirer pleinement parti des nouvelles possibilités et retombées offertes par les TIC. L'accessibilité aux données de la recherche est devenue une importante condition pour :

- Une gestion avisée de l'investissement public dans l'information factuelle ;
- Le développement de chaînes d'innovation à forte valeur ;
- L'accroissement de la valeur procurée par la coopération internationale.

Plus spécifiquement, l'amélioration de l'accès aux données et du partage de celles-ci :

- Renforce la liberté de l'investigation scientifique ;
- Encourage la diversité d'analyse et d'opinion ;
- Stimule de nouvelles recherches ;
- Rend possible l'expérimentation d'hypothèses et de méthodes d'analyse nouvelles ou différentes ;
- Contribue aux études sur les méthodes de collecte de données et leur mesure ;
- Facilite la formation des nouveaux chercheurs ;
- Permet l'exploration de thèmes qui n'avaient pas été envisagés par les chercheurs d'origine ;
- Conduit à la création de nouveaux ensembles de données par la combinaison de données provenant de multiples sources.

L'accès ouvert aux données de la recherche financée sur fonds publics et leur partage contribuent non seulement à maximiser l'impact des nouvelles technologies et des nouveaux réseaux numériques sur le potentiel de recherche, mais permettent aussi un retour plus important sur l'investissement public dans la recherche.

Dans tous les pays membres de l'OCDE, des volumes sans cesse croissants de données sont collectés par les chercheurs et établissements de recherche financés sur fonds publics. Cette somme de données de recherche en expansion rapide représente à la fois un investissement massif de fonds publics et l'une des sources potentielles des connaissances nécessaires pour relever les multiples défis auxquels est confrontée l'humanité.

Pour promouvoir un meilleur rendement scientifique et social des investissements publics dans les données de la recherche, les pays membres de l'OCDE ont établi diverses lois, politiques et pratiques sur l'accès aux données de la recherche au niveau national. Dans ce contexte, des lignes directrices internationales constitueraient un atout important pour encourager les échanges et l'utilisation à l'échelle mondiale des données de la recherche.

Les présents *Principes et Lignes directrices* visent les données de la recherche qui sont recueillies à l'aide de fonds publics pour produire des connaissances publiquement accessibles. La nature du « financement public » de la recherche varie considérablement d'un pays à l'autre, tout comme les politiques et pratiques actuelles en matière d'accès aux données en fonction des pays, des disciplines et des institutions. Ces différences appellent une approche flexible de l'accès aux données, et la reconnaissance du fait qu'il n'y a pas de solution unique. En outre, il appartiendra aux différents gouvernements nationaux et à leur milieu scientifique d'arbitrer entre les coûts d'un meilleur accès aux données de recherche et les bénéfices pouvant en résulter.

Indépendamment des différences qui peuvent exister entre les pratiques et politiques en matière de partage de données, ainsi que des restrictions légitimes auxquelles peut être soumis l'accès aux données, un partage plus systématique serait bénéfique pour pratiquement tous les types de recherche. Comme le relèvent les auteurs de l'étude de l'US National Research Council, intitulée *Bits of Power* :

> *La valeur des données réside dans leur exploitation. L'accès total et ouvert aux données scientifiques devrait devenir la norme internationale pour l'échange des données scientifiques issues de la recherche financée sur fonds publics.*

Les présents *Principes et Lignes directrices* ont pour finalités et objectifs spécifiques :

- De promouvoir une culture d'ouverture et de partage des données de la recherche entre les divers milieux de la recherche publique au sein des pays membres et à l'égard d'autres pays ;
- D'encourager l'échange de bonnes pratiques en matière d'accès et de partage des données ;
- De contribuer à une meilleure prise de conscience des coûts et bénéfices potentiels des restrictions et limitations actuelles visant l'accès aux données de la recherche financée sur fonds publics et le partage de ces données ;

- De mettre en évidence la nécessité de prendre en compte les réglementations et les pratiques en matière d'accès aux données et de partage des données dans l'élaboration des politiques et programmes scientifiques des pays membres ;

- De proposer un cadre de principes opérationnels, défini d'un commun accord, pour la mise en place de dispositifs d'accès aux données de la recherche dans les pays membres ;

- De fournir aux pays membres des recommandations sur les moyens d'améliorer le contexte international pour le partage et la diffusion des données de la recherche.

Les *Principes et Lignes directrices* exposés dans le présent document devraient aider les gouvernements, les organismes d'aide à la recherche et de financement de la recherche, les établissements de recherche et les chercheurs eux-mêmes à surmonter les obstacles et relever les défis liés à l'amélioration du partage des données de la recherche et de l'accès aux données, au plan international. Ces *Principes et Lignes directrices* doivent être interprétés et appliqués à la lumière des grandes questions que pose la fourniture de l'accès aux données :

- Questions technologiques : l'accès aux données de recherche et l'exploitation optimale de ces données nécessitent une infrastructure technologique adaptée, une large convergence de vues au plan international en matière d'interopérabilité, ainsi que des mécanismes efficaces de contrôle de la qualité des données.

- Questions institutionnelles et de gestion : bien qu'une meilleure accessibilité soit importante pour toutes les communautés scientifiques, la diversité des activités scientifiques donne à penser que des approches sur mesure et des modèles institutionnels variés de la gestion des données seront sans doute plus efficaces pour répondre aux besoins des chercheurs.

- Questions financières et budgétaires : l'infrastructure des données scientifiques nécessite une planification budgétaire spécifique continue et un soutien financier adéquat. L'exploitation des données de la recherche ne pourra être optimisée si les coûts d'accès, de gestion et de conservation doivent être imputés après coup aux projets de recherche. Il faut néanmoins souligner que les coûts de stockage et de gestion des données ont fortement diminué ces dernières années, et le manque d'informations sur ces changements peut en soi représenter un obstacle au progrès.

- Questions juridiques et de politiques : les législations nationales et les accords internationaux, notamment dans des domaines comme les droits de propriété intellectuelle et la protection de la vie privée, influent directement sur les pratiques en matière d'accès aux données et de partage des données et doivent être pleinement pris en compte dans l'élaboration des dispositifs d'accès aux données.

- Questions de culture et de comportement : la mise en place de mécanismes de formation et de rétribution adaptés est une composante nécessaire de la promotion des pratiques en matière d'accès et de partage des données. Ces considérations valent pour ceux qui financent, produisent, gèrent et exploitent les données de la recherche.

En oeuvrant pour améliorer l'accès aux données de la recherche dans le cadre de ces *Principes et Lignes directrices*, les pays membres devront arbitrer entre les coûts d'un meilleur accès à ces données et les bénéfices pouvant en découler. Les efforts déployés pour améliorer l'accès devront évidemment tenir compte des limitations financières existantes.

PRINCIPES ET LIGNES DIRECTRICES POUR L'ACCÈS AUX DONNÉES DE LA RECHERCHE FINANCÉE SUR FONDS PUBLICS

I. Objectifs

Ces *Principes et Lignes directrices pour l'accès aux données de la recherche financée sur fonds publics* (ci-après dénommés « *Principes et Lignes directrices* ») fournissent aux organismes publics d'élaboration de la politique scientifique et de financement des pays membres des recommandations générales sur l'accès aux données de la recherche financée sur fonds publics. Ils ont pour but de promouvoir l'accès aux données et leur mise en commun entre les chercheurs, établissements de recherche et organismes de recherche nationaux, tout en reconnaissant et en prenant en compte la diversité des lois, des politiques de recherche et structures administratives des pays membres.

L'objectif ultime des *Principes et Lignes directrices* est d'améliorer l'efficience et l'efficacité du système scientifique mondial, et non d'entraver son développement par de pesantes obligations et réglementations, ou d'imposer de nouveaux coûts aux systèmes scientifiques nationaux.

II. Champ d'application et définitions

Ces *Principes et Lignes directrices* visent les données de la recherche, déjà existantes ou à venir, qui sont subventionnées par des fonds publics pour produire des recherches et des connaissances scientifiques publiquement accessibles. Les *Principes et Lignes directrices* ne sont donc pas destinés à s'appliquer aux données de la recherche recueillies en vue de commercialiser les résultats de la recherche, ni aux données de la recherche qui appartiennent à une entité du secteur privé. L'accès à ces données fait intervenir tout un ensemble de considérations qui sortent du cadre du présent document. De plus, dans certains cas, l'accès aux données ou leur usage peut être limité afin de protéger la vie privée, la confidentialité ou des résultats couverts par un droit de propriété ou par la sécurité nationale.

Données de la recherche

Dans le cadre de ces *Principes et Lignes directrices*, les « données de la recherche » sont définies comme des enregistrements factuels (chiffres, textes, images et sons), qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires pour valider des résultats de recherche. Un ensemble de données de recherche constitue une représentation systématique et partielle du sujet faisant l'objet de la recherche.

Ce terme ne s'applique pas aux éléments suivants : carnets de laboratoire, analyses préliminaires et projets de documents scientifiques, programmes de travaux futurs, examens par les pairs, communications personnelles avec des collègues et objets matériels (par exemple, les échantillons de laboratoire, les souches bactériennes et les animaux de laboratoire tels que les souris). L'accès à tous ces produits ou résultats de la recherche est régi par d'autres considérations que celles abordées ici.

Ces Principes et Lignes directrices portent essentiellement sur les données de la recherche sur support numérique exploitable sur ordinateur. C'est en effet ce format qui offre le plus de possibilités d'améliorer la distribution efficiente des données et leur application pour la recherche, dans la mesure où les coûts marginaux de la transmission de données via l'internet sont pratiquement nuls. Les *Principes et Lignes directrices* pourraient également s'appliquer à des données de recherche sous forme analogique lorsque les coûts marginaux d'accès à ces données peuvent être maintenus à un niveau raisonnablement bas.

Données de la recherche financée sur fonds publics

Les données de la recherche financée sur fonds publics sont définies comme les données provenant de recherches menées par des organismes ou services publics ou à l'aide de fonds publics, quel que soit le niveau de gouvernement qui les fournit. La nature des « fonds publics » attribués pour la recherche étant très variable d'un pays à l'autre, ces *Principes et Lignes directrices* reconnaissent que ces différences plaident en faveur d'une approche souple de l'amélioration de l'accès aux données de la recherche.

Dispositifs d'accès

Les dispositifs d'accès sont définis comme étant le cadre réglementaire, stratégique et procédural mis en place par les établissements de recherche, les organismes de financement de la recherche et les autres acteurs concernés pour déterminer les conditions d'accès aux données de la recherche et d'utilisation de ces données.

III. Principes

A. Ouverture

Par ouverture, on entend l'accès dans des conditions d'égalité de la communauté scientifique internationale, à un coût le plus bas possible, de préférence ne dépassant pas le coût marginal de la diffusion. L'accès ouvert aux données de la recherche financée sur fonds publics devrait être aisé, rapide, convivial et passer de préférence par l'internet.

B. Flexibilité

La flexibilité suppose de prendre en compte les évolutions rapides et souvent imprévisibles des technologies de l'information, les caractéristiques de chaque domaine de recherche et la diversité des systèmes de recherche, des systèmes juridiques et des cultures de chaque pays membre. Les implications nationales, sociales, économiques et réglementaires spécifiques devraient être analysées lorsque des organisations conçoivent des dispositifs d'accès aux données de la recherche, et lorsque les gouvernements élaborent des politiques pour promouvoir l'accès aux données et examinent la mise en œuvre de ces *Principes et Lignes directrices*.

C. Transparence

L'information sur les données de la recherche et les organisations productrices de données, la documentation sur les données ainsi que les spécifications des conditions qui régissent leur utilisation devraient être accessibles au plan international, en toute transparence, dans l'idéal via l'internet. Le manque de visibilité des sources de données de recherche existantes et futures fait sérieusement obstacle à l'accès.

Les facteurs à considérer pour assurer la transparence sont notamment les suivants :

- L'information sur les organisations productrices de données ainsi que sur les données qu'elles détiennent, leur documentation concernant les ensembles de données disponibles et les conditions d'utilisation de ces données devraient pouvoir être aisément accessibles sur l'internet.
- Les organisations de recherche et les institutions publiques de recherche devraient activement diffuser l'information relative à leurs politiques en matière de données de la recherche auprès des chercheurs, des associations académiques, des universités et des autres acteurs de la recherche financée sur fonds publics.

- Chaque fois que cela peut être utile, tous les membres des divers milieux de la recherche devraient contribuer à établir des accords sur des normes de catalogage des données. L'application des normes existantes devrait être envisagée, dans la mesure du possible, pour éviter de solliciter davantage les ressources de la recherche et d'accroître la charge de travail des chercheurs et de leurs institutions.

- Des informations devraient être échangées sur la gestion des données et les conditions de leur accès entre les institutions chargées des archives de données et celles productrices de données, de manière à mettre en commun les pratiques exemplaires.

D. Conformité au droit

Les dispositifs d'accès aux données devraient respecter les droits et intérêts légitimes de tous les acteurs de l'activité de recherche publique.

L'accès à certaines données de la recherche et leur utilisation seront nécessairement limités par divers types de prescriptions légales, qui peuvent imposer des restrictions pour raisons de :

- Sécurité nationale : certaines données relatives au renseignement, aux activités militaires ou à la prise de décision politique peuvent être classifiées et partant, soumises à un accès limité.

- Protection de la vie privée et confidentialité : les données relatives aux sujets humains et d'autres données personnelles font l'objet d'un accès limité en vertu des législations et des politiques nationales de protection de la confidentialité et de la vie privée. Il convient toutefois que les détenteurs de ces données envisagent des procédures d'anonymisation ou de confidentialité permettant d'assurer un niveau de confidentialité satisfaisant afin de préserver autant que possible l'utilité des données pour les chercheurs.

- Secrets commerciaux et droits de propriété intellectuelle : les données concernant les entreprises ou autres parties, ou provenant de celles-ci, qui contiennent des informations confidentielles peuvent être inaccessibles pour la recherche.

- Protection d'espèces rares, menacées ou en danger : dans certains cas, il peut exister des raisons légitimes de limiter l'accès aux données sur la localisation de ressources biologiques en vue d'assurer leur préservation.

- Procédure légale : les données en cours d'examen dans le cadre de poursuites en justice (sub judice) peuvent être inaccessibles.

L'adhésion à des codes de conduite professionnels peut faciliter le respect des prescriptions légales.

E. Protection de la propriété intellectuelle

Les dispositifs d'accès aux données devraient tenir compte de l'applicabilité du droit d'auteur ou des autres législations sur la propriété intellectuelle pouvant concerner les bases de données de la recherche financée sur fonds publics. Les facteurs à prendre en compte sont les suivants :

- Compte tenu de la multiplication des partenariats public-privé pour le financement de la recherche et la collecte de données destinées à la recherche, des arrangements équilibrés public-privé devraient faciliter, le cas échéant, un large accès aux données de la recherche. La participation du secteur privé à la collecte des données ne devrait pas, en soi, être une raison pour restreindre l'accès aux données. Il conviendrait de prendre en considération des mesures qui favorisent l'accès et l'exploitation à des fins non commerciales tout en protégeant les intérêts commerciaux, comme la communication différée ou partielle des données ou l'adoption volontaire de mécanismes d'octroi de licences. De telles mesures peuvent permettre aux principaux participants d'exploiter pleinement les données de la recherche, sans en interdire inutilement l'accès.

- Dans les juridictions où les données et informations issues de la recherche publique sont protégées par des droits de propriété intellectuelle, les détenteurs de ces droits devraient néanmoins faciliter l'accès à ces données, en particulier pour la recherche publique ou pour d'autres fins répondant à l'intérêt général.

F. Responsabilité formelle

Les dispositifs d'accès devraient promouvoir des pratiques institutionnelles explicites et formalisées, telles que l'élaboration de règles et de réglementations, sur les responsabilités des diverses parties intervenant dans les activités liées aux données. Ces pratiques devraient concerner la paternité des données, la mention des producteurs, la propriété, la diffusion, les restrictions d'utilisation, les modalités financières, les règles éthiques, les conditions de licence, la responsabilité civile et l'archivage durable.

Les dispositifs d'accès, au niveau gouvernemental comme institutionnel, devraient être élaborés en consultation avec des représentants de tous les acteurs directement concernés. Dans les programmes ou projets de recherche en collaboration, et particulièrement dans le cadre de la coopération scientifique internationale ou de projets de recherche fondés sur des partenariats public-privé s'appuyant sur des cadres réglementaires différents, les parties

intéressées devraient négocier des accords de partage de données le plus tôt possible au cours du projet de recherche, et dans l'idéal pendant la phase de proposition initiale. On pourra ainsi veiller à ce qu'il soit tenu compte comme il convient et en temps opportun de questions telles que l'allocation de ressources pour le partage et la préservation durable des données de la recherche, les différences dans les législations nationales sur les droits de propriété intellectuelle, les limitations motivées par des considérations de sécurité nationale, ainsi que la protection de la vie privée et de la confidentialité.

Les dispositifs d'accès devraient également prendre en compte des facteurs tels que les caractéristiques des données, leur valeur potentielle pour la recherche, le niveau de traitement des données (données brutes, partiellement traitées ou finales), le fait qu'il s'agisse de données homogènes issues des instruments ou capteurs d'une installation ou de données hétérogènes collectées sur le terrain par des chercheurs isolés, de données sur des sujets humains ou de paramètres physiques, ou encore de données qui sont ou non générées directement par une entité gouvernementale ou grâce à un financement public. Ces variantes dans l'origine ou la nature des données devraient être prises en considération lors de l'établissement des dispositifs d'accès aux données.

De plus, il convient de prendre en compte les éléments suivants :

- Bon nombre de problèmes liés à l'accès, à la diffusion et au partage de données résultent d'un manque d'accords institutionnels explicites sur les modalités d'accès et d'utilisation. La gestion des données devenant toujours plus complexe dans certains domaines de recherche, les arrangements informels conclus traditionnellement entre les chercheurs risquent de ne plus être adaptés et de devoir être complétés par des pratiques et des procédures convenues en bonne et due forme.

- La responsabilité des divers aspects de l'accès aux données et de leur gestion devrait être établie dans les documents concernant les tâches officielles des instituts, les demandes de subventions, les contrats de recherche, les accords de publication et les licences, par exemple.

- La pérennité de l'infrastructure requise pour l'accès aux données revêt une importance particulière. Les établissements de recherche et les organismes publics devraient assumer officiellement la responsabilité de faire en sorte que les données de la recherche soient efficacement préservées, gérées et rendues accessibles de façon à pouvoir être exploitées de manière efficiente et adéquate sur le long terme.

G. Professionnalisme

Les dispositifs institutionnels pour la gestion des données de la recherche devraient être fondés sur les normes professionnelles applicables et sur les valeurs inscrites dans les codes de conduite des milieux scientifiques concernés.

Les facteurs à prendre en compte sont les suivants :

- L'utilisation de codes de conduite/de déontologie par les chercheurs et la communauté scientifique pourrait contribuer à simplifier et à réduire la réglementation régissant l'accès.
- La confiance mutuelle entre les chercheurs, ainsi qu'entre ces derniers et leurs institutions et autres organisations, joue un rôle important dans l'élaboration et le maintien de ces codes de conduite.
- Dans la pratique courante de la recherche, le chercheur ou l'institution qui produit initialement les données est parfois récompensé par l'utilisation exclusive temporaire des données. Les règles applicables à de tels arrangements incitatifs devraient être élaborées et explicitement précisées par les organismes sources de financement, en coopération avec les milieux de la recherche concernés.

Dans certains domaines scientifiques, le manque de préparation et de mise en œuvre de la documentation adéquate et de l'archivage des ensembles de données constitue l'un des obstacles majeurs à l'exploitation maximale de l'investissement dans les données de recherche. Lors de la planification des projets et programmes, à tous les niveaux, les problèmes de données devraient être abordés dès les tout premiers stades afin de prendre en compte les besoins en termes de fonds ou d'assistance technique pour les activités essentielles d'organisation et de conservation de ces ensembles de données. Il conviendrait de prêter attention aux incitations et au développement des compétences professionnelles dans tous les domaines de la gestion des données de recherche.

H. Interopérabilité

L'interopérabilité technologique et sémantique est essentielle pour faciliter et encourager l'accessibilité et l'utilisation des données de la recherche dans un contexte international et interdisciplinaire. Les dispositifs d'accès devraient tenir dûment compte des normes internationales pertinentes applicables en matière de documentation des données. Les pays membres et les établissements de recherche devraient coopérer avec des organisations internationales chargées de l'élaboration de nouvelles normes.

Bien que la science devienne une entreprise fortement mondialisée, l'incompatibilité des normes techniques et procédurales utilisées peut être un obstacle très sérieux aux multiples usages des ensembles de données.

Les facteurs qui devraient notamment être pris en compte sont les suivants :

- Il conviendrait de mentionner explicitement les normes employées dans la mesure où il s'agit de la première exigence de l'interopérabilité.
- Le fait de prendre pour modèle les disciplines les plus avancées à cet égard devrait être encouragé, en particulier par les organisations professionnelles internationales se consacrant à la science ainsi qu'à la collecte et à la conservation des données à des fins scientifiques et technologiques.
- Les travaux d'organisations chargées de définir des normes plus générales concernant les technologies de l'information et des communications devraient également être pris en considération.

I. Qualité

La valeur et l'utilité des données de recherche dépendent pour une large part de la qualité des données elles-mêmes. Les gestionnaires de données et les organisations de collecte de données devraient particulièrement veiller au respect de normes de qualité explicites. Lorsque ces normes n'existent pas encore, les institutions et les associations de recherche devraient, avec leurs chercheurs, se consacrer à leur élaboration. Si tous les domaines de la recherche peuvent tirer parti d'une meilleure qualité des données, certains exigent des normes beaucoup plus rigoureuses que d'autres. Pour cette seule raison, il n'est pas réaliste d'envisager des normes universelles de qualité des données. Les normes devraient être élaborées en consultation avec les chercheurs pour s'assurer que le niveau de qualité et de précision répond aux besoins des différentes disciplines.

Plus spécifiquement,

- Les dispositifs d'accès aux données devraient décrire les pratiques exemplaires quant aux méthodes, techniques et instruments employés pour le recueil, la diffusion et l'archivage accessible des données, afin de permettre un contrôle de qualité par un examen mutuel et d'autres moyens assurant la qualité et l'authenticité.
- L'origine des sources devrait être établie et spécifiée de façon vérifiable. Tous ceux qui souhaitent utiliser les données devraient pouvoir obtenir facilement ces informations, qu'il conviendrait

d'incorporer dans les métadonnées accompagnant les ensembles de données. Il est important de constituer ces métadonnées pour permettre aux scientifiques de comprendre les implications exactes des ensembles de données.

- Autant que possible, l'accès aux données devrait être lié à l'accès aux documents de recherche originaux, et les reproductions des ensembles de données devraient être liées aux originaux pour faciliter la validation des données et la détection d'erreurs dans ces ensembles.

- Les établissements de recherche et les associations professionnelles devraient concevoir des pratiques adéquates concernant les références de données et l'enregistrement de ces références sous forme d'index, dans la mesure où ce sont des indicateurs précieux de la qualité des données.

J. Sécurité

Il convient de s'attacher en particulier à encourager l'utilisation de techniques et d'instruments destinés à garantir l'intégrité et la sécurité des données de recherche. En ce qui concerne la garantie de l'intégrité d'un ensemble de données, tout devrait être mis en œuvre pour s'assurer du caractère complet des données et de l'absence d'erreurs. En ce qui concerne la sécurité, les données, de même que les métadonnées et descriptions correspondantes, devraient être protégées contre la perte, la destruction, la modification et l'accès non autorisé, intentionnels ou non, en conformité avec des protocoles de sécurité explicites. Les ensembles de données et les équipements servant à leur conservation devraient également être protégés contre les risques environnementaux tels que chaleur, poussières, surtensions, champs magnétiques et décharges électrostatiques.

K. Efficience

L'un des buts essentiels poursuivis en s'attachant à promouvoir l'accès et le partage des données est d'améliorer l'efficience globale de la recherche scientifique financée sur fonds publics afin d'éviter une duplication inutile et coûteuse des efforts de collecte de données.

Il convient de prendre en compte les aspects suivants :

- Les dispositifs d'accès aux données devraient favoriser l'amélioration du rapport coût-efficacité au sein du système scientifique mondial par la description des pratiques exemplaires dans les services de gestion de données et de soutien spécialisé.

- Si les données de la recherche financée sur fonds publics sont soumises à la règle par défaut d'ouverture énoncée en vertu du Principe A, cela ne signifie pas pour autant que toutes ces données doivent être conservées de manière permanente. Les acteurs de l'archivage de données devraient mener des évaluations coûts-avantages régulières et concevoir et affiner en permanence des protocoles visant à s'assurer que les ensembles de données ayant la plus grande utilité potentielle sont conservés et rendus accessibles. Le recours à des protocoles acceptés de conservation et à une information détaillée sur les données devrait contribuer à réduire les doubles emplois ainsi qu'à établir la sélectivité nécessaire en matière de conservation.

- Des services de soutien spécialisés, par exemple dans le cadre d'une collaboration sur des projets de recherche spécifiques avec des spécialistes hors de la sphère universitaire ou le recours à des organisations spécialisées dans la gestion des données, devraient être envisagés pour veiller à la rentabilité de la production, de l'utilisation, de la gestion et de l'archivage des données de la recherche.

- Si les chercheurs ou les producteurs de bases de données ne bénéficient pas d'incitations suffisantes, ils risquent de relâcher leurs efforts dans les activités liées aux données. Il conviendrait pour éviter ce problème d'envisager l'élaboration de nouvelles structures d'incitations et l'adaptation des structures existantes, notamment la prise en compte des activités de gestion de données dans les procédures de nomination et de promotion.

L. Responsabilité de rendre compte

Le fonctionnement des dispositifs d'accès aux données devrait faire l'objet d'une évaluation périodique par les groupes d'utilisateurs, les institutions responsables et les organismes de financement de la recherche. Même si chaque partie utilisera sans doute des critères d'évaluation quelque peu différents, la somme totale des résultats devrait donner une image détaillée de la valeur des données et des régimes d'accès aux données. Ces évaluations devraient contribuer à accroître le soutien en faveur du libre accès aux données par le milieu scientifique et la collectivité.

Il convient de prendre en compte les aspects suivants pour l'élaboration des critères d'évaluation :

- Les investissements publics globaux dans la production et la gestion des données de la recherche.
- Les performances de gestion des organismes de collecte et d'archivage de données.
- Le degré de réutilisation des ensembles de données existants.
- Les connaissances générées par le réemploi de données existantes.
- Le recours à des exercices de prospective ciblés afin de déterminer la nature et le champ des activités de préservation des données et les catégories de données les plus susceptibles d'être nécessaires à l'avenir.

Même s'il ne sera pas facile de se faire une idée claire et précise des coûts, des avantages et des résultats des dispositifs d'accès aux données, les responsables des dispositifs d'accès aux données devraient s'attacher à montrer les avantages d'un libre accès aux données afin de justifier et d'obtenir un soutien durable de la part de tous les niveaux de gouvernement.

M. Pérennité

Il conviendrait de tenir dûment compte de la pérennité de l'accès aux données de la recherche financée sur fonds publics comme l'un des éléments-clés des infrastructures de recherche. Cela suppose d'assumer la responsabilité administrative des mesures destinées à garantir un accès permanent aux données jugées comme devant être durablement conservées. La tâche peut être difficile étant donné que la plupart des projets de recherche et les financements publics accordés sont de durée limitée, alors que garantir l'accès aux données produites est une entreprise qui s'inscrit dans la durée. Les organismes de financement de la recherche et les établissements de recherche devraient donc étudier la préservation à long terme des données dès le début de chaque nouveau projet, et notamment rechercher les structures d'archivage les plus appropriées pour les données.

LES ÉDITIONS DE L'OCDE, 2, rue André-Pascal, 75775 PARIS CEDEX 16
IMPRIMÉ EN FRANCE
(00 2007 39 3 P) – n° 88180 2007